WE CLAIM:

Sub A1

~~1. A system for load balancing in a network environment comprising:~~

~~a plurality of servers coupled to a network;~~

~~a set of network resources associated with each of the servers, wherein at least some of the network resources are redundant;~~

~~a client coupled to the network and generating a request specifying some of the redundant resources;~~

~~a gateway machine coupled to the network in communication with the client, the gateway machine configured to receive the request from the client, select from amongst the servers that are associated with the request-specified redundant services, establish a communication channel with the selected server, and access the specified server to service the received client request; and~~

~~means coupled to the gateway machine for selecting amongst servers of redundant resources a particular server for a received request so as to balance load amongst the servers providing redundant resources.~~

~~2. The system of claim 1 further comprising means within the gateway machine for selecting amongst a plurality of channels between the gateway machine and the servers associated with the network resources specified by the client request.~~

~~3. The system of claim 1 wherein at least some of the plurality of servers comprise world wide web servers.~~

~~4. The system of claim 3 wherein the client comprises a web browser.~~

Sub A1

5. The system of claim 1 wherein each of the plurality of servers is associated with a network address and the gateway machine comprises a web server having an network address distinct from the plurality of servers.

6. The system of claim 1 wherein the network comprises the Internet.

7. The system of claim 1 further comprising means for monitoring quality of service between the gateway and each of the servers.

8. The system of claim 7 wherein the means for monitoring quality of service is implemented within the gateway machine.

9. The system of claim 7 further comprising means for selecting from amongst the servers providing redundant services using the relative quality of service between the servers.

10. The system of claim 7 further comprising:

means for allocating an additional server with redundant services in response to the quality of service falling below a preselected level, wherein the gateway machine is configured to establish a new communication channel through the network with the additional server.

11. The system of claim 1 wherein the gateway machine further comprises:

a means for generating a response to the client request using services provided to the gateway machine by the servers.

12. A method for load balancing in a network environment comprising:

providing a communication network;

Sub A'7

providing a plurality of servers coupled to the network wherein at least some of the servers provide redundant services;

generating a request for a redundant service in a network-coupled client machine;

directing the request to a network-coupled gateway machine;

causing the gateway machine to select amongst servers providing the redundant services a particular server for the received request so as to balance load amongst the plurality of servers providing the redundant services; and

after selecting a particular server, causing the gateway machine to generate a request to the selected server for the specified resources.

13 The method of claim 12 selecting amongst a plurality of channels between the gateway machine and the servers of redundant services specified by the client request.

14. The method of claim 12 wherein at least some of the plurality of servers comprise world wide web servers and the client comprises a web browser.

15. The method of claim 12 wherein the network comprises the Internet.

16. The method of claim 12 further comprising monitoring quality of service between the gateway and each of the servers.

17. The method of claim 16 further comprising selecting from amongst the servers providing redundant services using the relative quality of service between the servers.

Sub A^c 7

18. The method of claim 12 further comprising using the gateway machine to generate a response to the client request using services provided to the gateway machine by the servers.

19. A system for improving performance of a network connected server comprising:

an intermediary server receiving requests for server access from a plurality of sources;

means for monitoring at least one variable affecting server performance; and

a queue data structure within the intermediary server, wherein the queue data structure is responsive to the means for monitoring to release requests from the queue data structure in a manner determined to improve server performance.